

Full Brillouin-Zone, Charge Self-consistent Quantum Transport Simulations Enabled by Parallelization of the Nanoelectronic Modeling Tool (NEMO 1-D) on a Beowulf Cluster

Gerhard Klimeck

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

Tel.: (818) 354-2182 Fax.: (818) 393-3155

email: gekco@jpl.nasa.gov website: <http://hpc.jpl.nasa.gov/PEP/gekco>

As microelectronic research moves devices to nanometer scale operating at GHz speeds, the physics of electron flow through devices becomes more complicated and physical effects, which previously could be ignored safely, become significant. High energy electron injection, quantization of charge, quantization of energy, and electron scattering interactions are some of the phenomena that are presently being investigated experimentally and theoretically. TI/Raytheon developed a 1-D quantum device simulator (NEMO-1D) to address such issues. NEMO 1-D represents the state-of-the-art tool for quantitative transport simulations in resonant tunneling diodes. The work presented here is an extension of the NEMO 1-D software to parallel computing. This enabled, for the first time, the physical full band (sp³s* and sp³d⁵s* tight binding), full charge self-consistent simulation of high performance resonant tunneling diodes.

The availability of relatively cheap PC-based Beowulf clusters offers research and/or development groups an affordable entry into massively parallel computing. This work here describes the extension of NEMO 1-D to such parallel computers and discusses several different parallel algorithms reaching from coarse, to medium, to fine grain, and a mixed coarse/fine grain algorithm. The simulations are performed on 26 800MHz Pentium III Dual CPU nodes with 2GB/node RAM (52 CPUs). The nodes are connected by a high-end Myricom switch. The scaling of the various algorithms will be compared and the first full band, quantum charge self-consistent RTD simulation will be compared to experimental data.

NEMO's main task is the computation of current-voltage (I-V) characteristics for RTDs. The main workhorse model for high performance RTDs is the full band simulation, which is based on a numerical double integral of energy and transverse momentum over a transport kernel at each bias point. A full charge self-consistent simulation within this model on a single CPU is prohibitively expensive, as the generation of a single I-V would take about 1-2 weeks to compute. Simpler charge self-consistent models derived from a parabolic transverse subband assumption had been used in the past. Computation on a parallel computer now enables the thorough exploration of quantum mechanical transport including charge self-consistency effects in the whole Brillouin zone.

The simplest parallel algorithm for the computation of an I-V is the farming out of the different bias points to various CPUs and the final collection of the results. This type of algorithm is more efficient than any finer grain algorithms (due to its minimal communication requirements) and it minimizes the interference of parallelism with the 250,000 lines of code in NEMO 1-D. Such coarse grain algorithm was implemented and almost ideal wall clock time reduction with increasing number of CPU will be shown. However, this algorithm does have two drawbacks. 1) It is not useful for detailed studies at one or few bias points, and 2) it treats all bias points as independent (making charge feed-back from one bias point to the next impossible).

The double integral of energy and transverse momentum at each bias point offers two other opportunities for parallelism. Parallelism in the loop around all transverse momentum nodes (medium grain) and all energy nodes (fine grain) has been implemented and the resulting reduction of wall clock time with increasing number of CPUs will be discussed. Scaling for two different network speeds (slow 100Mb/s and fast 2Gb/s) will be discussed.

Finally a comparison of a fullband, quantum charge self-consistent simulation of a InGaAs/InAlAs RTD to experimental data will be presented for the first time. The parallelization of the NEMO 1-D code enabled the quantitative simulation of a complete RTD test matrix within a few days, while the experimental creation of such a test matrix took several weeks. The test matrix considered here consists of 4 different RTD structures simulated at 300K and 77K.

This work shows the utility of low-cost, high performance Linux clusters for the design and characterization of electronic devices using physics-based simulation software.